# Prediction of the insulin sensitivity index using Bayesian networks

Bøttcher, Susanne Gammelgaard; Dethlefsen, Claus

*Publication date:*
2006

*Document Version*
Publisher's PDF, also known as Version of record

# Prediction of the Insulin Sensitivity Index using Bayesian Networks

Susanne Gammelgaard Bøttcher
Aalborg University

Claus Dethlefsen
Aalborg Hospital

# Prediction of the insulin sensitivity index

- The insulin sensitivity index ($S_I$) can be used in assessing the risk of developing type 2 diabetes.
- Is determined by observations from an intravenous glucose tolerance test using Bergmans minimal model and estimated by a non-linear least squares estimation technique.
- Goal is to determine this index from an oral glucose tolerance test, to be used in large scale epidemiological studies.
- Idea: Learn Bayesian network with observations from an oral test and the index determined from an intravenous test.
- As learning tool, use the software package deal, written for R.
- Export the final network to Hugin to calculate the predictive distribution of $S_I$.

# Bayesian Network

We define it as:

- A Directed Acyclic Graph (DAG) $D = (V, E)$.
- Set of variables associated with D is $X = (X_v)_{v \in V}$.
- To each vertex $v$ with parents $pa(v)$, there is attached a local probability distribution, $p(x_v | x_{pa(v)})$.
- Possible lack of edges in D encodes conditional independencies,

$$p(x) = \prod_{v \in V} p(x_v | x_{pa(v)}).$$

# Parameter learning

- Continuous variables are Gaussian linear regressions on the continuous parents, and have therefore jointly a Gaussian distribution.
- Parameter learning

$$p(\theta|data) \propto p(data|\theta)p(\theta).$$

- Assume parameter independence and complete data. Posterior parameter independence follows.
- Conjugate prior distribution Gaussian - inverse Wishart.

# Structure Learning

- Structure learning

$$\mathrm{p}(\mathrm{DAG}|\mathrm{data}) \propto \mathrm{p}(\mathrm{data}|\mathrm{DAG})\mathrm{p}(\mathrm{DAG}).$$

- If all DAGs equally likely

$$\mathrm{p}(\mathrm{DAG}|\mathrm{data}) \propto \mathrm{p}(\mathrm{data}|\mathrm{DAG}).$$

- Network score decomposable.
- The likelihood is given as

$$\mathrm{p}(\mathrm{data}|\mathrm{DAG}) = \int_{\theta} \mathrm{p}(\mathrm{data}|\theta, \mathrm{DAG})\mathrm{p}(\theta|\mathrm{DAG})\mathrm{d}\theta.$$

- Automated procedure for finding $\mathrm{p}(\theta|\mathrm{DAG})$ for all possible DAGs.

# Master Prior Procedure

(Geiger and Heckerman (1994))

- Random variables have joint Gaussian distribution

$$(\mathbf{y}|\mathbf{m}, \Sigma) \sim \mathcal{N}(\mathbf{m}, \Sigma).$$

- Let joint parameter prior be Gaussian-inverse Wishart,

$$(\mathbf{m}|\Sigma) \sim \mathcal{N}(\mu, \frac{1}{\nu}\Sigma) \text{ and } (\Sigma) \sim \mathcal{IW}(\rho, \Phi).$$

  By marginalizing and conditioning, parameters for each parent-child relationship can be found.

- Parameters are independent, so joint prior is the product of the local priors.

- To specify master prior $\mathcal{N}(\mu, \frac{1}{\nu}\Sigma)$ and $\mathrm{IW}(\rho, \Phi)$, specify prior network $(\mathrm{D}, \mathcal{P})$ and sample size of imaginary database.

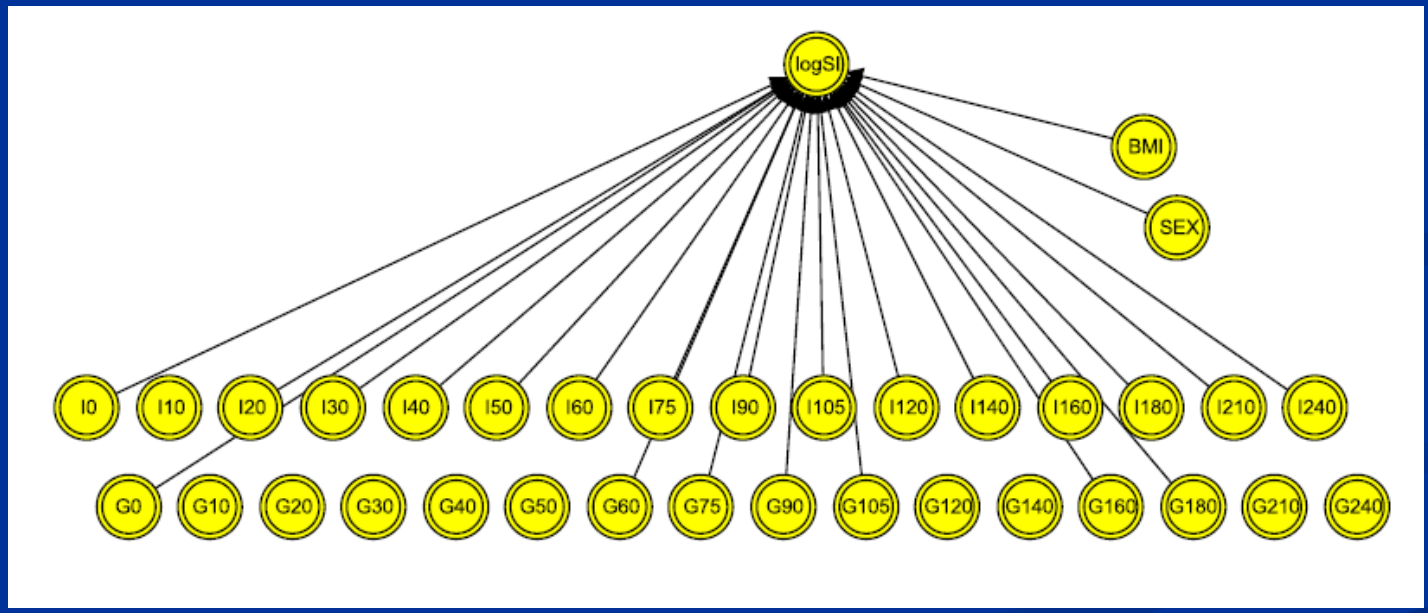
# Software: deal

- `deal`: A package for R.
- Downloads from CRAN `http://cran.R-project.org/`.
- Graphical interface.
- Local master prior procedure for determining parameter priors for all networks.
- Parameter learning.
- Calculation of network score.
- Structure learning using greedy search (with random restarts).
- Simulation of datasets.

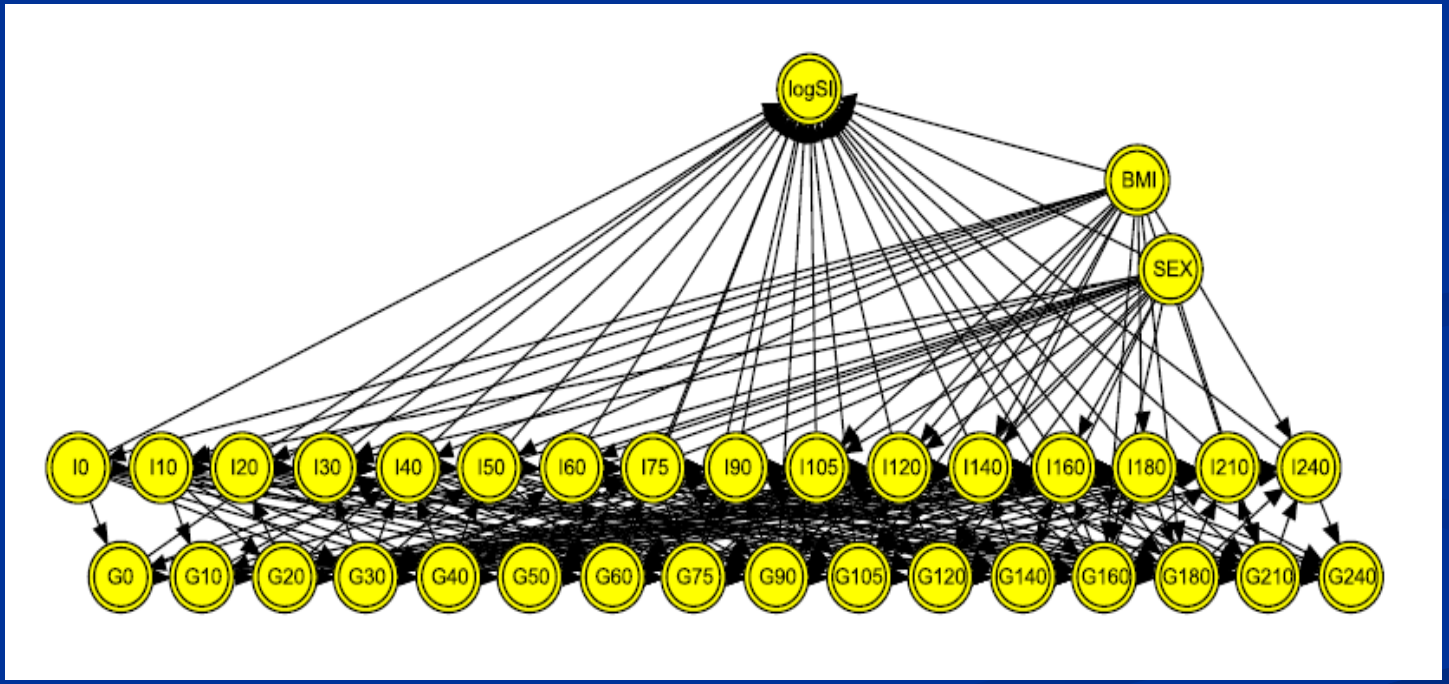
# Data

- 187 non-diabetic glucose tolerant subjects underwent both an OGTT and an IVGTT.
- 140 subjects used as training data and 47 subjects used as validation data.
- $\log S_I$ for each subject calculated from the IVGTT using Bergmans minimal model.
- Glucose and insulin concentrations determined in the OGTT at 0, 10, 20, 30, 40, 50, 60, 75, 90, 105, 120, 140, 160, 180, 210 and 240 minutes.
- Other variables BMI and sex. Sex modelled as a continuous variable.

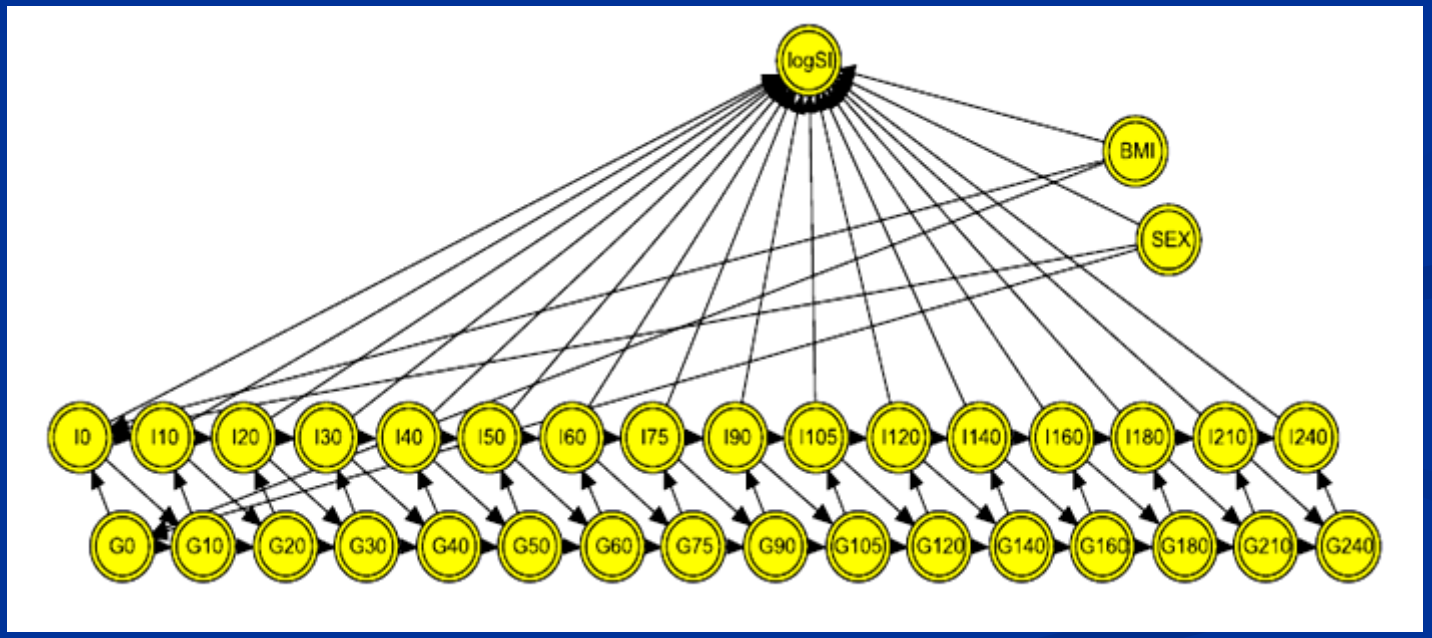
# Bayesian Regression Network



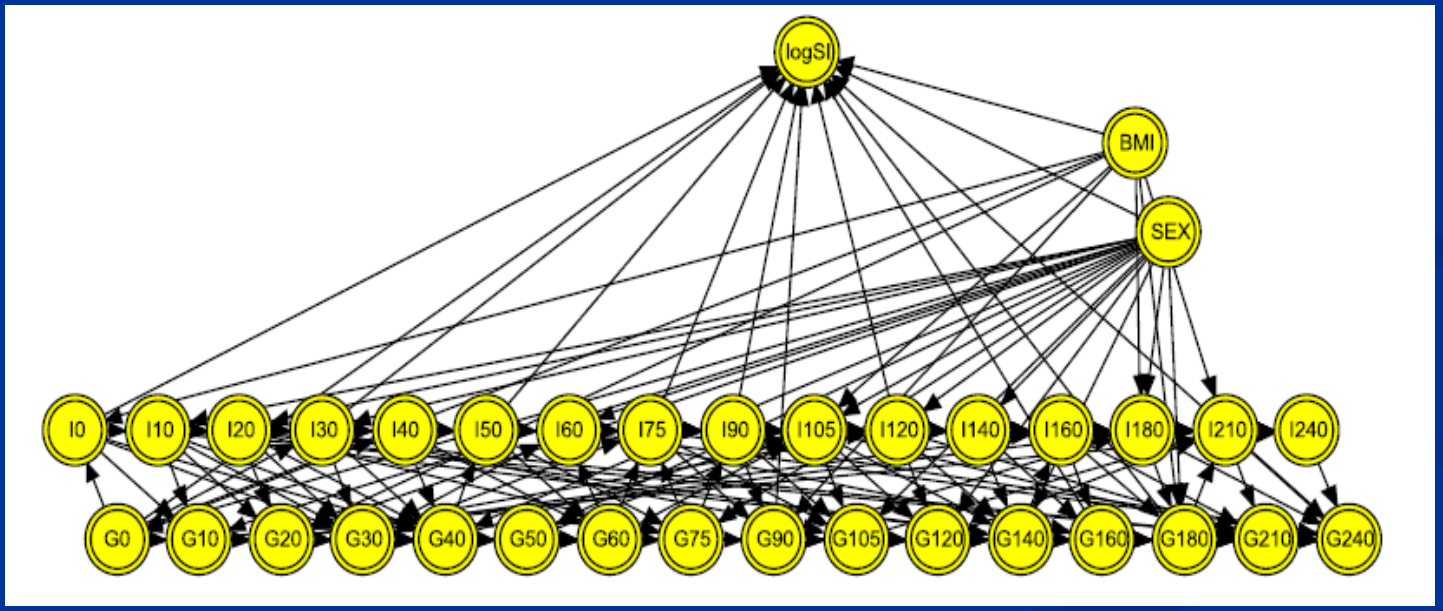
# Bayesian Network with empty prior



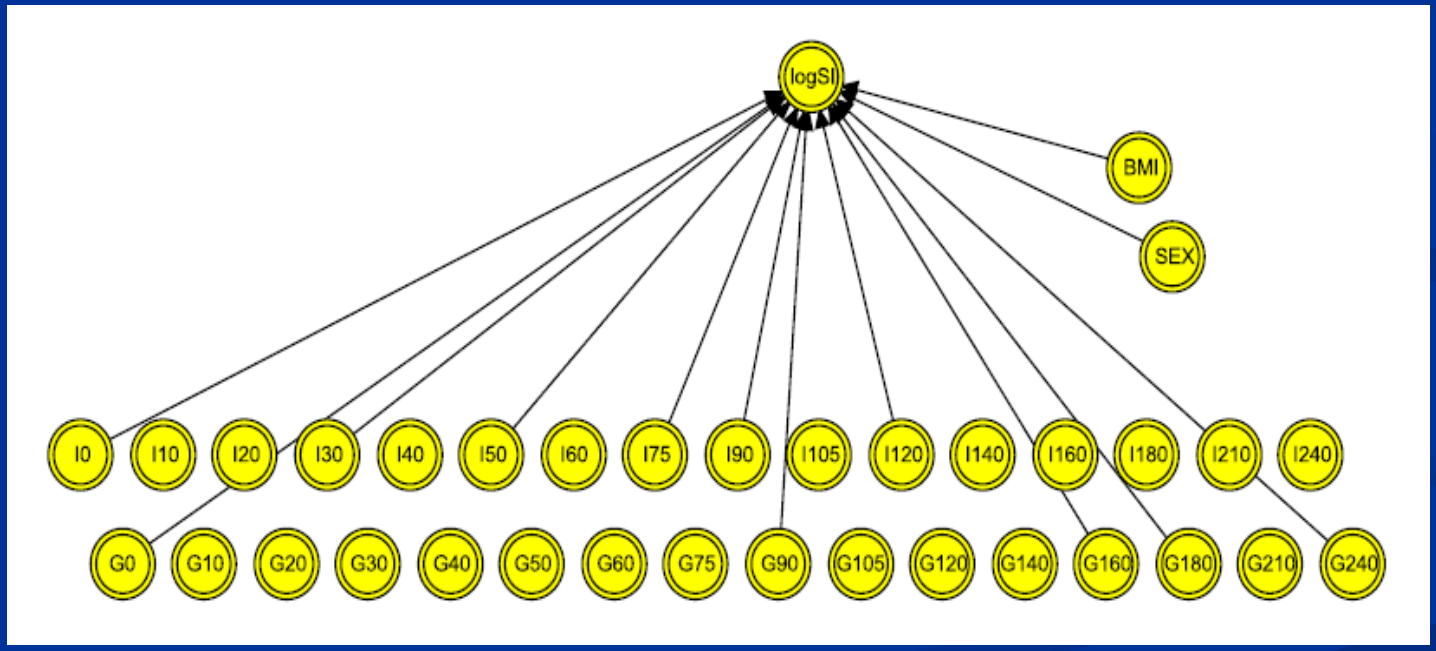
# The physiological Network



# Bayesian Network with physiological prior



# Markov Blanket with physiological prior



# Results using multiple linear regression

Previous result (Drivsholm et al. (2003)) using multiple linear regression:

$$\log S_I \sim \quad BMI + SEX + G0 + I0 + G30 + I30 + G60 + I60 + G105 + I105 + G180 + I180 + G240 + I240$$

Result using leaps and bound algorithm for best subset selection:

$$\log S_I \sim \quad BMI + SEX + I50 + I90 + G160$$

# Evaluation

- Network scores (log scores).
- For each subject the conditional distribution of $\log S_I$ is calculated given the observations from the OGTT using Hugin (the predictive distribution of $\log S_I$).
- 95%'s credibility intervals $\mu \pm 1.96 \cdot \sigma$ for det predictive distribution of $\log S_I$ is calculated. Examine whether 95 $\%$ of the corresponding IVGTT $\log S_I$ values lie within this interval. If so, the predictive distribution of $\log S_I$ is *well calibrated*.
- Residual standard deviation (SD) and correlation coefficient, ($R^2$) obtained from a linear regression of the IVGTT obtained $\log S_I$ on the predicted $\log S_I$. To see if there is systematic bias in these regressions, the intercept and slope of these regressions lines are reported.

# Network scores

| Model | Empty prior | Physiological prior |
|---|---|---|
| BR | −17878.30 | −17848.33 |
| BN | −16528.39 | −14851.44 |
| MLR | −17886.17 | −17849.06 |
| L&B | −17894.95 | −17846.12 |

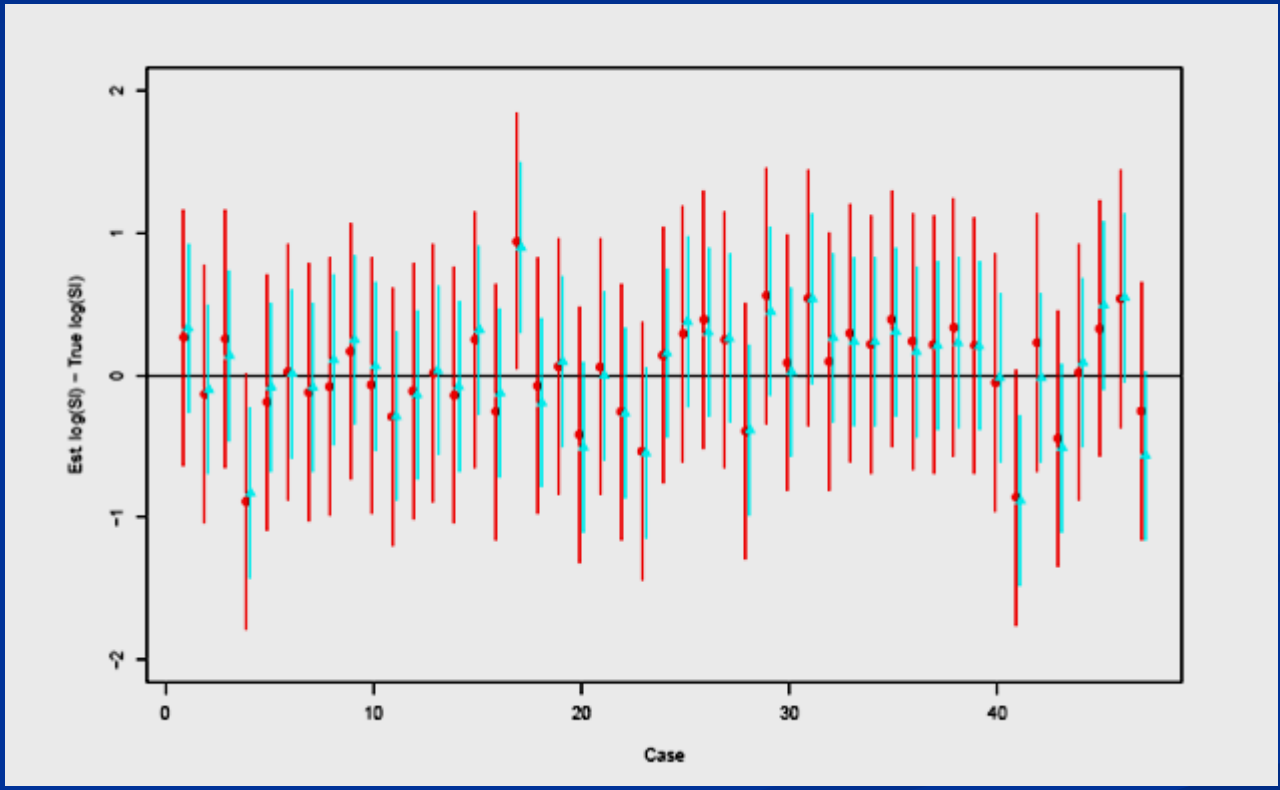
# Intercept and slope for regression of IVGTT obtained index on predicted index

| Model | Tr. data (intercept, slope) | Val. data (intercept, slope) |
|---|---|---|
| BR empty prior | $(-0.19, 1.09)$ | $(-0.05, 1.01)$ |
| BN empty prior | $(-0.03, 1.01)$ | $(0.25, 0.87)$ |
| BN phsy. prior | $(-0.06, 1.03)$ | $(0.14, 0.92)$ |
| MLR | $(0, 1)$ | $(0.06, 0.96)$ |
| L&B | $(0, 1)$ | $(0.11, 0.93)$ |

# $R^2$ and SD from the regression and also number of values outside credibility intervals.

| Model | Tr. data $R^2$ (SD) | Val. data $R^2$ (SD) | Tr. outside | Val. outside |
|---|---|---|---|---|
| BR with empty prior | 0.76(0.31) | 0.73(0.35) | 1(1%) | 1(2%) |
| BN with empty prior | 0.76(0.31) | 0.73(0.35) | 1(1%) | 1(2%) |
| BN with phys. prior | 0.77(0.30) | 0.73(0.36) | 7(5%) | 3(6%) |
| MLR | 0.76(0.31) | 0.66(0.40) | 3(2%) | 3(6%) |
| L&B | 0.75(0.31) | 0.73(0.36) | 6(4%) | 4(9%) |

# Predicted index and credibility intervals – empty prior red and disks, physiological prior green and triangles



# Summary

- All approaches give adequate predictions of $S_I$.
- Bayesian network with physiological prior estimates most precise predictive distribution of $S_I$.
- Can use any prior knowledge available from e.g. previous studies or the physiological understanding of the problem.
- Predictive distribution of $S_I$ can be calculated in situations with missing observations - here the correlations between the insulin and glucose observations are important.

# References


Bergman, R. N., Ider, Y. Z., Bowden, C. R. and Cobelli, C. (1979). Quantitative estimation of insulin sensitivity, *American Journal of Physiology* **236**: E667–E677.

Bøttcher, S. G. (2001). Learning Bayesian Networks with Mixed Variables, *Artificial Intelligence and Statistics 2001*, Morgan Kaufmann, San Francisco, CA, USA, pp. 149–156.

Bøttcher, S. G. and Dethlefsen, C. (2003). `deal`: A Package for Learning Bayesian Networks, *Journal of Statistical Software* **8**(20): 1–40.

Cooper, G. and Herskovits, E. (1992). A Bayesian method for the induction of probabilistic networks from data, *Machine Learning* **9**: 309–347.

Cowell, R. G., Dawid, A. P., Lauritzen, S. L. and Spiegelhalter, D. J. (1999). *Probabilistic Networks and Expert Systems*, Springer-Verlag, Berlin-Heidelberg-New York.

# References


Dawid, A. P. (1982). The Well-Calibrated Bayesian, *Journal of the American Statistical Association* **77**(379): 605–610.

Drivsholm, T., Hansen, T., Urhammer, S. A., Palacios, R. T., Vølund, A., Borch-Johnsen, K. and Pedersen, O. B. (2003). Assessment of insulin sensitivity index and acute insulin reponse from an oral glucose tolerance test in subjects with normal glucose tolerance, Novo Nordisk A/S.

Furnival, G. M. and Wilson, R. W. (1974). Regression by Leaps and Bounds, *Technometrics* **16**(4): 499–511.

Geiger, D. and Heckerman, D. (1994). Learning Gaussian Networks, *Proceedings of Tenth Conference on Uncertainty in Artificial Intelligence*, Morgan Kaufmann, San Francisco, CA, USA, pp. 235–243.

Haughton, D. M. A. (1988). On The Choice of a Model to fit Data From an Exponential Family, *The Annals of Statistics* **16**(1): 342–355.

Heckerman, D., Geiger, D. and Chickering, D. M. (1995). Learning Bayesian networks: The combination of knowledge and statistical data, *Machine Learning* **20**: 197–243.

# References


Ihaka, R. and Gentleman, R. (1996). R: A language for data analysis and graphics, *Journal of Computational and Graphical Statistics* **5**: 299–314.

Lauritzen, S. L. (1992). Propagation of probabilities, means and variances in mixed graphical association models, *Journal of the American Statistical Association* **87**(420): 1098–1108.

Martin, B. C., Warram, J. H., Krolewski, A. S., Bergman, R. N., Soeldner, J. S. and Kahn, C. R. (1992). Role of glucose and insulin resistance in development of type 2 diabetes mellitus: results of a 25-year follow-up study, *The Lancet* **340**: 925–929.

Pacini, G. and Bergman, R. N. (1986). MINMOD: a computer program to calculate insulin sensitivity and pancreatic responsivity from the frequently sampled intraveneous glucose tolerance test, *Computer Methods and Programs in Biomedicine* **23**: 113–122.